

Weekly Report

Done

1. PVis 出结果分数除了最后一位以外都很低。攻击点出在了这些方面：
 - a. 相关工作由于做过压缩所以篇幅较少。他们都说可以多应用游戏相关的可视化工作，尽管我们心里很清楚这些工作跟我们的确实不算很相关；
 - b. Case study 嫌这些不真实不信服
 - c. 不是很信用户真的有从中得到了 advantage (相比于应用简单的可视化工具)
 - d. 模型还是太复杂看不懂 数据模型篇幅大而无聊 为什么没用别的模型？
 - e. 仍有人质疑 generalizability 的问题
2. 与邓老师一些讨论以及我的思考
 - a. 利用陆地交通数据可视化结果做预测(比如天气或者其他变量)，再迁移到其他交通比如船只交通数据. 或者其他可视化结果拿来预测的方法
 - b. Graph layout 的迁移. 这个由于对 graph layout 了解不多，并不知道是否可以在不同 layout 之间迁移？需要简单调研。此外我还能想到的可能是不同网络类型，比如社交网络以及服务器网络之间进行某种 layout 可视化结果的迁移。
 - i. 基本不会有从小数据到大数据的迁移
 - ii. 对于 graph 可视化结果的”好”与”不好”，比如以上述社交网络与计算机网络为例，我如何知道一个好的 layout，是否与分析任务有关？
 - c. Visual saliency 的问题，我给他展示了去年那篇文章\并觉得可以套用马哥的思路，他表示感兴趣并且也想做但是没有进一步阐述
 - d. 他觉得可以直接把马哥的 scatterplot 升级到 scatterplot matrix (SPLOM) 上做一个类似的工作。
 - i. 主旨就是在高维数据中找一些相关性或者其他 interesting 的东西(就是 subspace)
 - ii. 有些情况下可以通过调整维度顺序(如果我们找到一种方法，通过调换顺序使得用户更容易发现有意思的模式，也是很好的)；但有些情况维度顺序需要保留，这时需要外面套 RNN 这样的带时间序列的模型
3. 与南电的一些讨论
4. 准备组会报告时，我想到之前有调研过一些 provenance visualization 的工作；但是这些多为 provenance 的可视化，但这些 provenance 作为用户使用和探索记录的日志，其实也是可以作推荐的。
 - a. Provenance 可以包括用户交互的记录\背后数据的流动\可视化结果的记录等等.
 - b. 以上面 SPLOM 为例，用户在探索高维的 SPLOM 可能常常有点丢失方向，如果我们记录了 provenance (交互\数据\甚至图片，也即是可视化结果)，那我们就可以通过不同的方式去给用户推荐可能有趣的维度，既可以基于交互，也可以看数据的覆盖率(我已经交互的看了哪些还有哪些可能有趣)，而图片则是从另一种角度来分析信息比如借用马哥的 ScatterNet 来找到相似的或者不相似的.
 - c. 另外，如果有及时反馈，也考虑套用强化学习的模型来为可视化探索提供推荐；这里的关键问题是需要一个简单但又能有及时反馈的可视化；并且需要大量数据.
5. 修改大黑书中的图片
6. 找了一些前几年 vis 文章，试图下周做一些 formalization 的工作

7. 找到了一本很不错的理论与实践结合的教材(<hands on machine learning with scikit-learn and tensorflow>), 让小顾和课程招生来夯实基础和学习一些新的知识;我也在学一些新的并复习之前的 tensorflow 相关编程.

To Do

1. 想就上面 2,4 两点继续讨论.
2. 完成 6 中论文的简单整理, 看看有没有思路
3. 找老师讨论 BeXplorer 该何去何从

论文阅读

ScatterNet: A Deep Subjective Similarity Model for Visual Analysis of Scatterplots 在生成数据上面非常讲究, 最大化覆盖的原则, hard example 的选取等, 顺便复习了之前一些图像处理用的 NN 的知识如 triplet loss 等等. 文章对于”相似”这种度量的选取也很巧妙; 但在 SPLOM 中, 所谓 interesting pattern 已经不是用相似这么简单可以描述的了.

Visual Exploration of Large Scatter Plot Matrices by Pattern Recommendation based on eye tracking 本文就定义了一下他们认为的 interesting: 未探索的或者与已经探索的非常 dissimilar 的. 然后他们用 eye tracking 来了解用户已经干了什么. 这个思路非常清奇, 我甚至也觉得可以用在 Graph 探索的推荐上.

在上面提到的教材中我学习了 autoencoder 背后的 intuition 以及从最简单版本的线性 autoencoder 到最近有点火的变分 autoencoder. 以及强化学习的入门阶段的知识.

以及上述 6 中的论文筛选浏览了一些文章.

工作时间

平时 9, 周末一共 8 小时. 总共 53.

个人规划

性质	Ddl	进展	目标
中长期	2019 年中	/	统计 优化 算法
短期	18 年底	一些讨论和新的脑洞, 有很多还是跟之前的一些探索有点沾边, 会继续讨论与了解	XX 学习对可视化支持
长期			博士毕业论文中的研究方向